

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-91191

(43)公開日 平成10年(1998) 4月10日

(51)Int.Cl.⁶

G 1 0 L 5/04

識別記号

F I

G 1 0 L 5/04

F

審査請求 未請求 請求項の数6 O L (全 12 頁)

(21)出願番号

特願平8-246718

(22)出願日

平成8年(1996) 9月18日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 鏡嶋 岳彦

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(72)発明者 赤嶺 政巳

神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

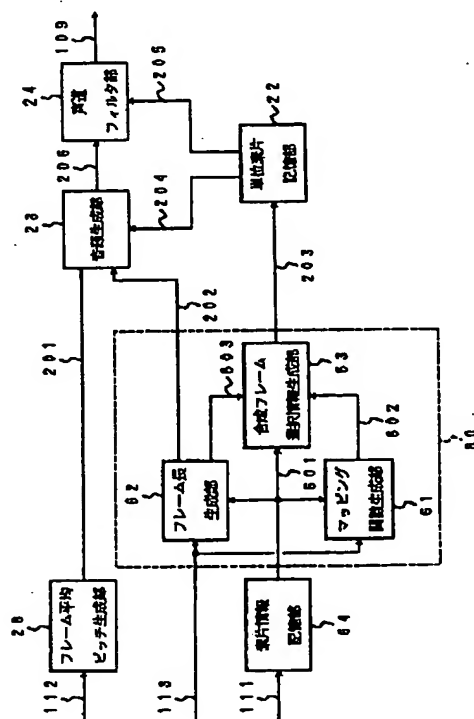
(74)代理人 弁理士 鈴江 武彦 (外6名)

(54)【発明の名称】 音声合成方法

(57)【要約】

【課題】自然性を損なうことなく任意の音韻継続時間長の合成音声を生じ得る音声合成方法を提供する。

【解決手段】合成しようとする音声信号の性質を表す第1のフレーム単位の合成パラメータ(音源パラメータ204、声道パラメータ205)を用いて音声信号109を合成する音声合成方法において、単位素片記憶部22に合成パラメータとなる第2のフレーム単位の単位素片を記憶しておき、入力された音韻継続時間長113の情報に従ってマッピング関数生成部61で生成されたマッピング関数を合成フレーム選択情報生成部63に与えて、第1のフレームと第2のフレームとの対応付けを行い、かつマッピング関数に基づいて第1のフレームに対応する第2のフレームの単位素片を間引くかまたは繰り返すことにより、合成パラメータを生成する。



【特許請求の範囲】

【請求項 1】合成しようとする音声信号の性質を表す第 1 のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、

前記合成パラメータとなる第 2 のフレーム単位の単位素片を記憶しておき、

入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第 1 のフレームと第 2 のフレームとの対応付けを行い、

第 1 のフレームに対応する第 2 のフレームの単位素片を前記合成パラメータとして生成することを特徴とする音声合成方法。

【請求項 2】合成しようとする音声信号の性質を表す第 1 のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、

前記合成パラメータとなる第 2 のフレーム単位の単位素片を記憶しておき、

入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第 1 のフレームと第 2 のフレームとの対応付けを行い、

かつ該マッピング関数に基づいて第 1 のフレームに対応する第 2 のフレームの単位素片を間引くかまたは繰り返すことにより、前記合成パラメータを生成することを特徴とする音声合成方法。

【請求項 3】合成しようとする音声信号の性質を表す第 1 のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、

前記合成パラメータとなる第 2 のフレーム単位の単位素片を記憶しておき、

入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第 1 のフレームと第 2 のフレームとの対応付けを行うとともに、

該マッピング関数に基づいて第 1 のフレームのフレーム長を変化させ、

第 1 のフレームに対応する第 2 のフレームの単位素片を前記合成パラメータとして生成することを特徴とする音声合成方法。

【請求項 4】合成しようとする音声信号の性質を表す第 1 のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、

前記合成パラメータとなる第 2 のフレーム単位の単位素片を記憶しておき、

入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第 1 のフレームと第 2 のフレームとの対応付けを行い、

前記単位素片の素片長が前記音韻継続時間長より小さい場合は該マッピング関数に基づいて第 1 のフレームに対応する第 2 のフレームの単位素片を間引くかまたは繰り返すことにより前記合成パラメータを生成し、

前記単位素片の素片長が前記音韻継続時間長以上の場合

には該マッピング関数に基づいて第 1 のフレームのフレーム長を変化させ、

第 1 のフレームに対応する第 2 のフレームの単位素片を前記合成パラメータとして生成することを特徴とする音声合成方法。

【請求項 5】前記マッピング関数として、その傾きが音韻境界付近で予め定められた略一定の値となる関数を用いることを特徴とする請求項 1～4 のいずれか 1 項に記載の音声合成方法。

【請求項 6】前記マッピング関数を前記単位素片の過渡部と定常部を指定する情報および前記音韻継続時間長とに従って生成することを特徴とする請求項 1～4 のいずれか 1 項に記載の音声合成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、テキスト音声合成のための音声合成方法に係り、特に音韻記号列、ピッチおよび音韻継続時間長などの情報から音声信号を生成する音声合成方法に関する。

【0002】

【従来の技術】任意の文章から人工的に音声信号を作り出すことをテキスト音声合成という。テキスト音声合成システムは、一般的に図 8 に示されるように、言語処理部 1、音韻処理部 2 および音声合成器 3 の 3 つの要素から構成される。入力されたテキスト 115 は、まず言語処理部 1 において形態素解析や構文解析などが行われ、次に韻律処理部 2 においてアクセントやイントネーションの処理が行われて、音韻記号列 111、ピッチパターン 112 および音韻継続時間長 113 などの情報が出力される。最後に、音声合成器 3 で音節、音素、1 ピッチ区間などの基本となる小さな単位（単位素片）の特徴パラメータを音韻記号列、ピッチパターン、音韻継続時間長などの情報に従って選択し、ピッチや継続時間長を制御して接続することによって合成音声信号 109 を生成する。

【0003】自然音声では、発話の速度、意味内容、前後の音韻などの影響で、音韻の継続時間長が様々に変化している。そこで、自然音声に近い高品質な音声を作成するためには、任意の継続時間長の音声を合成器によって合成することが可能でなければならない。

【0004】このような継続時間長を変化させることができる音声合成器として、特開昭 61-290499

「発声速度可変の音声合成装置」が知られている。この音声合成器の原理は、音声の定常部において単位素片のフレームを間引いて合成フレーム列を構成するとともに、合成フレームのフレーム長を一様に変化させることによって継続時間長を変化させるものである。

【0005】この従来の音声合成器の構成について図 9 を用いて説明する。図 9 において、フレーム平均ピッチ生成部 28 は、ピッチパターン 112 よりフレーム平均

ピッチ周期201を求めて出力する。素片情報記憶部25は、単位素片フレームのフレーム数、定常部の範囲および過渡部の範囲などの素片情報を記憶しており、入力された音韻記号列111に従って、対応する素片の素片情報207を出力する。フレーム長生成部26は、素片情報207に基づいて継続時間長が音韻継続時間長113と等しくなるようにフレーム長202と定常部伸縮率208を求めて出力する。合成フレーム選択情報生成部27は、素片情報207と定常部伸縮率208に従って合成フレームのフレーム選択情報203を出力する。

【0006】単位素片記憶部22は、例えば合成単位を音素とすると、全ての母音および全ての子音に対応する単位素片の情報を記憶している。各単位素片は、例えば当該音韻の音声信号を一定のフレーム周期で分析することによって生成され、複数の単位素片フレームにより構成される。各単位素片フレームは音源パラメータおよび声道パラメータから構成されており、フレーム選択情報

$$(v n_a + n_b) f = T$$

合成フレーム情報生成部27では、単位素片フレームのうち定常部のフレームについて、定常部伸縮率208に従って間引きまたは繰返しを行って合成フレーム列を生成し、各フレームのフレーム選択情報203を出力する。このようにして、任意の音韻継続時間長の音声を作成することが可能である。

【0010】

【発明が解決しようとする課題】上述した従来の音声合成器では、音声の定常部の間引きと過渡部の伸縮によって音韻継続時間長の変化を実現しているが、自然音声においては必ずしもこのようなモデルで音韻継続時間の異なる音声を表すことはできない。すなわち、従来の音声合成器は明瞭性を保つには有効であるが、そもそも自然音声では定常部と過渡部の境界は明確ではなく、性質が常に滑らかに変化しているため、過渡部と定常部に分類して異なる処理を行うことは、自然性が低下する原因となる。

【0011】また、自然音声を分析することによって単位素片を生成する場合、合成単位のフレームを過渡部と定常部に分類する必要があるが、このような分類は視察によって行わざるを得ないため、多大な労力を必要とする。

【0012】本発明は、このような従来の問題点を解決するためになされたもので、自然性を損なうことなく任意の音韻継続時間長の合成音声を生じ得る音声合成方法を提供することを目的とする。

【0013】

【課題を解決するための手段】上記の課題を解決するため、本発明は合成しようとする音声信号の性質を表す第1のフレーム単位の合成パラメータを用いて音声信号を合成するに際し、合成パラメータとなる第2のフレーム単位の単位素片を記憶しておき、入力された音韻継続時間

203に従って選択されたフレームの音源パラメータ204と声道パラメータ205が合成フレームの合成パラメータとして出力される。

【0007】音源生成部23は、フレーム平均ピッチ周期201と音源パラメータ204に従って、フレーム長202の長さの音源信号206を出力する。声道フィルタ部24は、声道パラメータ205によって特性が決定される声道フィルタを音源信号206で駆動することにより合成音声信号109を生成する。

【0008】フレーム長生成部26の詳細な動作について説明する。素片情報207より得られる当該単位素片の定常部フレーム数および過渡部フレーム数をそれぞれ n_a 、 n_b とし、音韻継続時間長113を T とすると、合成フレーム長202 (f) と定常部伸縮率208 (v) は、次式を満たすように決定される。

【0009】

(1)

間長の情報に従って生成されたマッピング関数に基づいて第1のフレームと第2のフレームとの対応付けを行い、第1のフレームに対応する第2のフレームの単位素片を合成パラメータとして生成することの特徴とする。

【0014】ここで、「合成パラメータ」とは、声道特性、音源特性、1ピッチ長の音声波形など、音声信号の性質を表すパラメータを表している。また、子音を C 、母音を V でそれぞれ表すと、「単位素片」とは、 C 、 V 、 CV 、 VCV 、 CVC などの合成単位に対応して生成された合成パラメータの系列を表すものとする。

【0015】本発明の第1の態様では、第1のフレームと第2のフレームとの対応付けを行った後、マッピング関数に基づいて第1のフレームに対応する第2のフレームの単位素片を間引くかまたは繰返すことにより、合成パラメータを生成する。

【0016】第1のフレームと第2のフレームとの対応付けを行うマッピング関数は、音韻継続時間長に基づいて生成されるため、このマッピング関数に基づいて第1のフレームに対応する第2のフレームの単位素片を間引くかまたは繰返すことで合成パラメータを生成すれば、合成音声信号の継続時間長が変化する。

【0017】本発明の第2の態様では、第1のフレームと第2のフレームとの対応付けを行うとともに、マッピング関数に基づいて第1のフレームのフレーム長を変化させ、第1のフレームに対応する第2のフレームの単位素片を合成パラメータとして生成する。

【0018】このようにマッピング関数に基づいて第1のフレームのフレーム長を変化させることによって、合成音声信号の継続時間長が変化する。本発明の第3の態様では、第1および第2の態様を組み合わせ、第1のフレームと第2のフレームとの対応付けを行った後、単位素片の素片長が音韻継続時間長より小さい場合は第1

のフレームに対応する第2のフレームの単位素片を間引くかまたは繰り返すことにより合成パラメータを生成し、単位素片の素片長が音韻継続時間長以上の場合にはマッピング関数に基づいて第1のフレームのフレーム長を設定し、第1のフレームに対応する第2のフレームの単位素片を合成パラメータとして生成する。

【0019】このように本発明によると、音声そのものでなく合成パラメータを変化させることで音韻時間長を変化させるため、マッピング関数として例えば傾きが音韻境界付近で予め定められた略一定の値となる関数を用いたり、マッピング関数を単位素片の過渡部と定常部を指定する情報および音韻継続時間長とに従って生成することにより、単位素片の伸縮度合いが連続的に変化することになるため、音韻継続時間長の変化による合成音声の自然性の低下がなく、音質の良好な音声合成が可能となる。

【0020】

【発明の実施の形態】以下では、音韻(C, V)を合成単位とする規則合成器に本発明を適用した場合の実施形態について説明するが、本発明はこのような合成単位に限定されるものではなく、その他の合成単位の場合も適用することが可能である。

【0021】(第1の実施形態)図1に、本発明の第1の実施形態に係る音声合成器の構成を示す。本実施形態は、合成しようとする音声信号の性質を表す第1のフレーム(合成フレーム)単位の合成パラメータを用いて音声信号を生成する場合、合成フレームのフレーム長を一定とし、合成フレームとなる第2のフレーム(単位素片フレーム)の間引きまたは繰り返しによって、合成音声の音韻継続時間長を変更するようにしたものである。

【0022】本実施形態において、フレーム平均ピッチ生成部28、単位素片記憶部22、音源生成部23および声道フィルタ部24の動作は、図9に示した従来の音声合成器と同様である。すなわち、フレーム平均ピッチ生成部28は、ピッチパターン112よりフレーム平均ピッチ周期201を求めて出力する。

【0023】単位素片記憶部22は、例えば合成単位を音素とすると、全ての母音および全ての子音に対応する単位素片の情報を記憶している。各単位素片は、例えば

単位素片フレームのフレーム数	: n'
合成フレームのフレーム数	: n
単位素片フレームのフレーム長	: f'
合成フレームのフレーム長	: f
単位素片長	: $T' = n' \cdot f'$
音韻継続時間長	: T
合成フレーム	: $F_i, (i=1, 2, 3, \dots, n)$
単位素片フレーム	: $F'_j, (j=1, 2, 3, \dots, n')$

フレーム長生成部62は、合成フレームのフレーム数 n およびフレーム長 f を次式に従って求める。

当該音韻の音声信号を一定のフレーム周期で分析することによって生成され、複数のフレーム(以下、単位素片フレームという)単位で構成される。単位素片フレームは、音源パラメータおよび声道パラメータから構成されており、単位素片記憶部22はフレーム選択情報203に従って選択された単位素片フレームの音源パラメータ204と声道パラメータ205を出力する。

【0024】音源生成部23は、フレーム平均ピッチ周期201と音源パラメータ204に従って、フレーム長202の長さの音源信号206を出力する。声道フィルタ部24は、声道パラメータ205によって特性が決定される声道フィルタを音源信号206で駆動することにより、合成音声信号109を生成する。

【0025】また、素片情報記憶部64は、入力された音韻記号列111に従って対応する素片の素片情報601を出力するが、この素片情報601には単位素片フレームのフレーム数、単位素片フレームのフレーム長などの情報が含まれ、定常部の範囲と過渡部の範囲の情報は必ずしも含まれない点で従来の音声合成器と異なっている。

【0026】次に、本実施形態の特徴的な構成である時間長制御部60について説明する。時間長制御部60は、マッピング関数生成部61、フレーム長生成部62および合成フレーム選択情報生成部63から構成されている。マッピング関数生成部61は、素片情報601と音韻継続時間長113に従って、合成フレーム列に対応する時間の変数から単位素片のフレームに対応する時間の変数へのマッピング関数602を生成する。フレーム長生成部62は、音韻継続時間長113と素片情報601に従って、合成フレームのフレーム数603およびフレーム長202を求める。合成フレーム選択情報生成部63は、フレーム数がフレーム数603で表される合成フレームと、フレーム数がフレーム情報601で表される単位素片フレームをマッピング関数602を用いて対応付けることによって合成フレーム列を生成し、各合成フレームのフレーム選択情報203を出力する。

【0027】以下、各構成要素の詳細な説明に当たり、記号を次のように定義する。

【0028】

【数1】

$$n = \left\lceil \frac{T}{f'} \right\rceil \quad (2)$$

$$f = T/n \quad (3)$$

【0029】ただし、記号 $\lceil \cdot \rceil$ は整数に丸める演算を表すものとする。マッピング関数生成部 61 は、合成フレーム列に対応する時間の変数 t から単位素片のフレームに対応する時間の変数 t' へのマッピング関数 $t' = m(t)$ を生成する。マッピング関数 $m(t)$ は、定義域 $m(0) = 0$

$$m(T) = T' \quad (5)$$

$$\frac{d}{dt} m(t) = m'(t) \geq 0 \quad (0 \leq t \leq T) \quad (6)$$

【0031】式 (6) は、関数 $m(t)$ が単調増加関数であることを表している。 $m(t)$ の導関数 $m'(t)$ は、マッピング関数 $m(t)$ の傾きに対応するため、単位素片の伸縮の度合を表している。すなわち、 $m'(t) > 1$ の区間では単位素片のフレームが間引かれ、

が $0 < t < T$ 、値域が $0 < m(t) < T'$ の関数である。また、マッピング関数 $m(t)$ は以下の条件式を満たさなければならない。

【0030】

$$\text{【数2】} \quad (4)$$

$m'(t) < 1$ の区間では単位素片のフレームが繰り返される。式 (4) (5) (6) の条件を満たす関数の例を以下に示す。

【0032】

【数3】

$$m_1(t) = \begin{cases} \frac{2(T-T')}{T^3} t^3 - \frac{3(T-T')}{T^2} t^2 + t & (3T' \geq T) \\ \frac{4}{27T'^2} t^3 - \frac{2}{3T'^2} t^2 + t & (3T' < T, t \leq \frac{3T}{2}) \\ \frac{T'}{2} & (3T' < T, \frac{3T}{2} < t < T - \frac{3T}{2}) \\ \frac{4}{27T'^2} (t-3T')^3 - \frac{2}{3T'^2} (t-3T')^2 + (t-3T') & (3T' < T, T - \frac{3T}{2} \leq t) \end{cases} \quad (7)$$

【0033】 $m_1(t)$ の導関数 $m'_1(t)$ は、 $0 \leq t \leq T$ において常に 0 以上となり、式 (6) を満たして

$$m'_1(0) = 1 \quad (8)$$

$$m'_1(T) = 1 \quad (9)$$

この式は、音韻の境界では単位素片を伸縮させないことを表している。すなわち、式 (7) のマッピング関数 $m_1(t)$ は、音韻の境界付近では単位素片の伸縮の度合が小さく、音韻の中央部では伸縮の度合が大きく、かつ伸縮の度合が連続的に変化するような関数である。

【0034】合成フレーム選択情報生成部 63 は、合成フレーム F_i 、($i=1, 2, 3, \dots, n$) と単位素片フレーム F'_j 、($j=1, 2, 3, \dots, n'$) をマッ

いる。また、 $m'_1(t)$ には次のような特徴がある。

ピング関数 $m(t)$ を用いて対応付ける。合成フレームと単位素片フレームの対応付けは、各合成フレームのフレーム中心時刻を $m(t)$ によってマッピングし、その時刻が含まれる単位素片フレームとそれぞれ対応付けることによって行う。

【0035】

【数4】

$$F_i = F'_{ji} \quad (10)$$

$$j_i = \left\lceil \frac{m(f(1-0.5))}{f'} \right\rceil \quad (11)$$

ただし、記号「 \lceil 」は整数に切り上げる演算を数す

【0036】次に、本実施形態による合成フレーム列の生成例を以下のパラメータの場合について、図2を用いて説明する。

単位素片フレーム数： $n' = 12$

単位素片フレーム長： $f' = 10 \text{ msec}$

単位素片長： $T' = n' f' = 120 \text{ msec}$

音韻継続時間長： $T = 45 \text{ msec}$

式(7)に $T = 45$ 、 $T' = 120$ を代入することによって、マッピング関数 $m'_1(t)$ は以下のように求められる。

【0037】

【数5】

$$m_1(t) = -\frac{2}{1215} t^3 + \frac{1}{9} t^2 + t \quad (12)$$

次に、合成フレーム数および合成フレーム長は、式

(2) (3) より以下のように求められる。

【0038】

【数6】

$$n = \left\lceil \frac{45}{10} \right\rceil = 5 \quad (13)$$

$$f = 45/5 = 9 \quad (14)$$

式(11)を用いてフレームをマッピングする。

【数7】

【0039】

$$\lceil m_1(9(1-0.5))/10 \rceil = 1 \quad (15)$$

$$\lceil m_1(9(2-0.5))/10 \rceil = 3 \quad (16)$$

$$\lceil m_1(9(3-0.5))/10 \rceil = 6 \quad (17)$$

$$\lceil m_1(9(4-0.5))/10 \rceil = 10 \quad (18)$$

$$\lceil m_1(9(5-0.5))/10 \rceil = 12 \quad (19)$$

【0040】以上の結果から、 $F_1 = F'_1$ 、 $F_2 = F'_3$ 、 $F_3 = F'_6$ 、 $F_4 = F'_{10}$ 、 $F_5 = F'_{12}$ となる。また、本実施形態の変形として、音韻の種類によって異なる種類のマッピング関数を用いることも可能である。例えば、閉鎖区間（無音区間）に続く破裂区間（有音区間）によって構成される無声破裂音などの場合、なるべく閉鎖区間で伸縮させることが望ましいため、後ろほど伸縮の度合いが小さくなるようなマッピング関数を用いればよい。

【0041】（第2の実施形態）図3に、本発明の第2の実施形態に係る音声合成器の構成を示す。本実施形態は、単位素片フレームの間引き／繰り返しは行わず、合成フレームのフレーム長を変化させることによって、合成音声の継続時間長を変化させるようにしたものである。すなわち、合成フレーム数 n は単位素片フレームのフレーム数 n' と等しく、単位素片フレームはそのまま合成フレーム列に対応する。

【0042】本実施形態において、フレーム平均ピッチ生成部28、単位素片記憶部22、音源生成部23、声道フィルタ部24および素片情報記憶部64の動作は、図1に示した第1の実施形態の場合と同様であるため、同一の参照符号を付して説明を省略する。

【0043】本実施形態の特徴的な構成である時間長制御部70について説明する。まず、マッピング関数生成部71は、素片情報601と音韻継続時間長113に従って、単位素片のフレームに対応する時間の変数から合成フレーム列に対応する時間の変数へのマッピング関数701を生成する。フレーム長生成部72は、マッピング関数701と素片情報601に従って、合成フレームのフレーム長202を求める。合成フレーム選択情報生成部73は、単位素片フレームのフレーム選択情報203を順次出力する。

【0044】以下、各構成要素の詳細な説明に当たり、記号を次のように定義する。

単位素片フレームのフレーム数: n'
 合成フレームのフレーム数: n
 単位素片フレームのフレーム長: f'
 合成フレームのフレーム長: $f_i, (i=1, 2, 3, \dots, n)$
 単位素片長: $T' = n' f'$
 音韻継続時間長: T
 合成フレーム: $F_i, (i=1, 2, 3, \dots, n)$
 単位素片のフレーム: $F'_j, (j=1, 2, 3, \dots, n)$

ただし、本実施形態においては単位素片フレームの間引

き／繰り返しを行わないため、次式の関係が成立する。

$$n = n' \quad (20)$$

$$F_i = F'_i \quad (i=1, 2, \dots, n) \quad (21)$$

マッピング関数生成部71は、単位素片フレームに対応する時間の変数 t' から合成フレーム列に対応する時間の変数 t へのマッピング関数 $t = r(t')$ を生成する。マッピング関数 $r(t')$ は、定義域が $0 < t' < T'$

T' , 値域が $0 < r(t') < T$ の関数である。また、 $r(t')$ は以下の条件式を満たさなければならない。

【0045】

【数8】

(22)

$$r(T') = T \quad (23)$$

$$\frac{d}{dt} r(t') = r'(t') \geq 0 \quad (0 \leq t' \leq T') \quad (24)$$

【0046】式(24)は $r(t')$ が単調増加関数であることを表している。 $r(t')$ の導関数 $r'(t')$ は、マッピング関数 $r(t')$ の傾きに対応するため、単位素片の伸縮の割合を表している。すなわち、 $r'(t') < 1$ の区間では合成フレームのフレーム長が単位素片フレームのフレーム長よりも短くなり、

$r'(t') > 1$ の区間では逆に長くなる。式(22)(23)(24)の条件を満たす関数の例を以下に示す。

【0047】

【数9】

$$r_1(t') = \begin{cases} \frac{2(T'-T)}{T^3} t'^3 - \frac{3(T'-T)}{T^2} t'^2 + t' & (3T \geq T') \\ \frac{4}{27T^2} t'^3 - \frac{2}{3T^2} t'^2 + t' & (3T < T', t' \leq \frac{3T}{2}) \\ \frac{T}{2} & (3T < T', \frac{3T}{2} < t' < T - \frac{3T}{2}) \\ \frac{4}{27T^2} (t'-3T)^3 - \frac{2}{3T^2} (t'-3T)^2 + (t'-3T) & (3T < T', T - \frac{3T}{2} \leq t') \end{cases}$$

(25)

【0048】 $r_1(t')$ の導関数 $r'_1(t')$ は、 $0 \leq t' \leq T$ において常に0以上となり、式(24)を満たしている。また、 $r'_1(t')$ には次のような特徴

$$r'_1(0) = 1$$

$$r'_1(T) = 1$$

この式は、音韻の境界では単位素片を伸縮させないことを表している。すなわち、式(25)のマッピング関数は、音韻の境界付近では単位素片の伸縮の割合が小さく、音韻の中央部では伸縮の割合が大きく、かつ伸縮の

がある。

【0049】

(26)

(27)

割合が連続的に変化するような関数である。

【0050】フレーム長生成部72は、合成フレームのフレーム長 $f_i, (i=1, 2, \dots, n)$ をマッピング関数 $r(t')$ を用いて生成する。合成フレームのフレ

ーム長 f_i は、次式より求められる。

$$f_i = r(f' \cdot i) - r(f' \cdot (i-1)) \quad (i=1, 2, \dots, n) \quad (28)$$

次に、本実施形態による合成フレーム列の生成例を以下のパラメータの場合について、図4を用いて説明する。ただし、図2との比較のため図4は横軸を t 、縦軸を

$$\begin{aligned} \text{単位素片フレームのフレーム数} : n' &= 5 \\ \text{単位素片フレームのフレーム長} : f' &= 10 \text{ msec} \\ \text{単位素片長} : T' &= n' \cdot f' = 50 \text{ msec} \\ \text{音韻継続時間長} : T &= 120 \text{ msec} \end{aligned}$$

式(25)に $T=120$ 、 $T'=50$ を代入することにより、マッピング関数は以下のように求められる。

$$r_1(t) = -\frac{7}{6250} t^3 + \frac{21}{250} t^2 + t \quad (29)$$

【0054】次に、合成フレーム長は式(28)より以

$$f_1 = 17.28 [\text{msec}] \quad (30)$$

$$f_2 = 27.36 [\text{msec}] \quad (31)$$

$$f_3 = 30.72 [\text{msec}] \quad (32)$$

$$f_4 = 27.36 [\text{msec}] \quad (33)$$

$$f_5 = 17.28 [\text{msec}] \quad (34)$$

また、本実施形態の変形として、定常部の範囲と過渡部の範囲の情報を素片情報に含めて、定常部の範囲と過渡部の範囲の情報と音韻継続時間長よりマッピング関数を求めることも可能である。マッピング関数としては、例えば図5で示されるような、過渡部で傾きが常に一定となり、定常部で傾きが滑らかに変化するような関数 $t = r_2(t')$ を用いることができる。

【0055】(第3の実施形態) 図6に、本発明の第3の実施形態に係る音声合成器の構成を示す。本実施形態は、単位素片の素片長が継続時間長よりも小さい場合は第1の実施形態と同様に動作し、単位素片の素片長が継続時間長以上の場合は第2の実施形態と同様に動作するものである。

【0056】本実施形態において、フレーム平均ピッチ生成部28、単位素片記憶部22、音源生成部23、声道フィルタ部24、素片情報記憶部64、時間長制御部60および時間長制御部70の動作は、図1に示した第1の実施形態および図3に示した第2の実施形態の場合と同様であるため、同一の参照符号を付して説明を省略する。

【0057】モード選択部81は、素片情報601と音韻継続時間長113より、単位素片の素片長と音韻継続時間長を比較してモード選択情報801を出力する。単位素片の素片長を T' とし、音韻継続時間長を T とすると、このモード選択情報801に従って、 $T > T'$ の場合には時間長制御部60が選択され、 $T \leq T'$ の場合には時間長制御部70が選択される。

【0058】第1の実施形態では、 $T < T'$ の場合にフレーム間引きによる情報欠落の可能性があり、また第2

t' として描いている。

【0052】

【0053】

【数10】

下のように求められる。

の実施形態では、 $T > T'$ の場合にフレーム長が長くなり過ぎてピッチ周期の変化が粗くなる可能性があるが、本実施形態では、第1および第2の実施形態を組み合わせることにより、情報欠落やピッチ周期の変化が粗くなることなく、より良好な合成音声を得られる。

【0059】(第4の実施形態) 図7に、本発明の第4の実施形態に係る音声合成器の構成を示す。本実施形態は、第1の実施形態におけるマッピング関数生成部61をマッピング関数選択部81およびマッピング関数記憶部82で置き換えたものである。

【0060】本実施形態において、フレーム平均ピッチ生成部28、単位素片記憶部22、音源生成部23、声道フィルタ部24、素片情報記憶部64、フレーム長生成部62、合成フレーム選択情報生成部63の動作は、図1に示した第1の実施形態の場合と同様であるため、同一の参照符号を付して説明を省略する。

【0061】マッピング関数記憶部82は、複数のマッピング関数を記憶しており、マッピング関数選択情報801に従って選択されたマッピング関数602を出力する。マッピング関数選択部81は、マッピング関数記憶部82に記憶されているマッピング関数の中から、素片情報601と音韻継続時間長113に従って、適したマッピング関数を選択し、マッピング関数選択情報801を出力する。

【0062】

【発明の効果】以上説明したように、本発明によれば合成パラメータとなる第2のフレーム単位の単位素片を記憶しておき、入力された音韻継続時間長の情報に従って

生成されたマッピング関数に基づいて第1のフレームと第2のフレームとの対応付けを行い、第1のフレームに対応する第2のフレームの単位素片を合成パラメータとして生成して、この合成パラメータを用いて音声信号を合成することにより、任意の継続時間長の音声合成することが可能であって、しかも単位素片の伸縮の度合を連続的に変化させることができるため、自然性にすぐれた合成音声を得ることができる。

【図面の簡単な説明】

【図1】 本発明の第1の実施形態を示すブロック図

【図2】 本発明の第1の実施形態における合成フレーム情報生成の一例を表す模式図

【図3】 本発明の第2の実施形態を示すブロック図

【図4】 第2の実施形態における合成フレーム情報生成の一例を表す模式図

【図5】 第2の実施形態における合成フレーム情報生成の他の例を表す模式図

【図6】 本発明の第3の実施形態を示すブロック図

【図7】 本発明の第4の実施形態を示すブロック図

【図8】 テキスト音声合成システムの構成を表すブロック図

【図9】 従来の音声合成器を示すブロック図

【符号の説明】

1…言語処理部
2…韻律処理部

3…合成器

61, 71…マッピング関数生成部

22…単位素片記憶部

23…音源生成部

24…声道フィルタ部

25, 64…素片情報記憶部

26, 62, 72…フレーム長生成部

27, 63, 73…合成フレーム選択情報生成部

28…フレーム平均ピッチ生成部

60, 70, 80…時間長制御部

81…マッピング関数選択部

82…マッピング関数記憶部

109…合成音声信号

111…音韻記号列

112…ピッチパターン

113…音韻継続時間長

201…フレーム平均ピッチ

202…合成フレームのフレーム長

203…合成フレームのフレーム選択情報

204…音源パラメータ

205…声道パラメータ

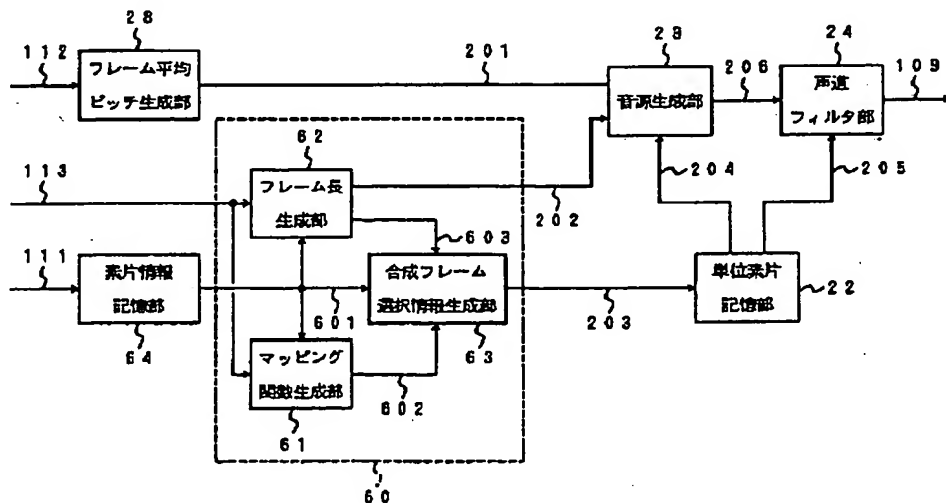
206…音源信号

601…素片情報

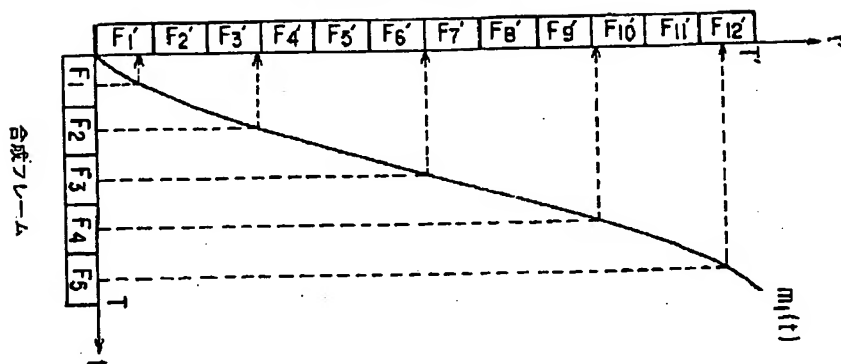
602…マッピング関数

603…合成フレームのフレーム数

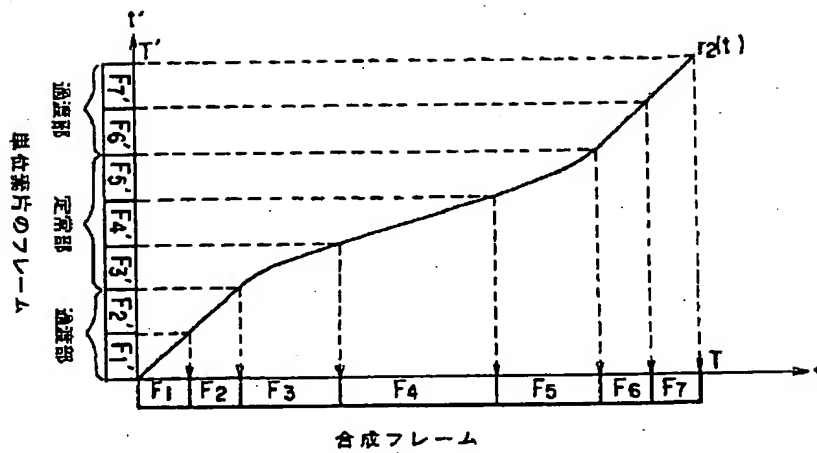
【図1】



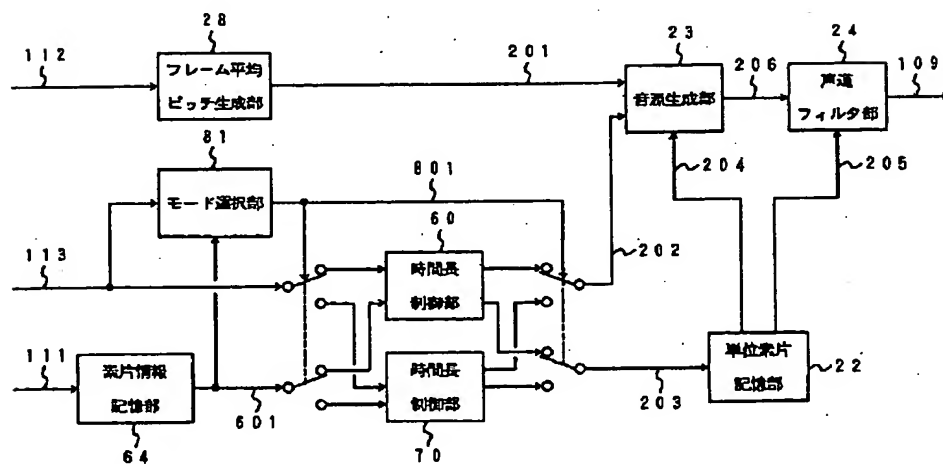
単位素片のフレーム



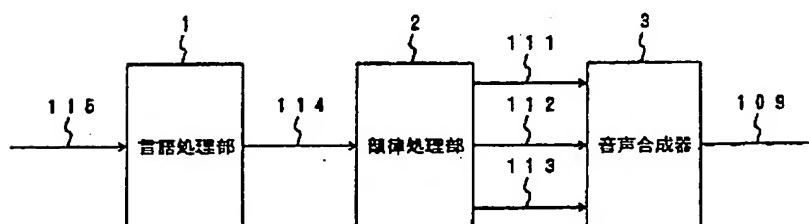
【図5】



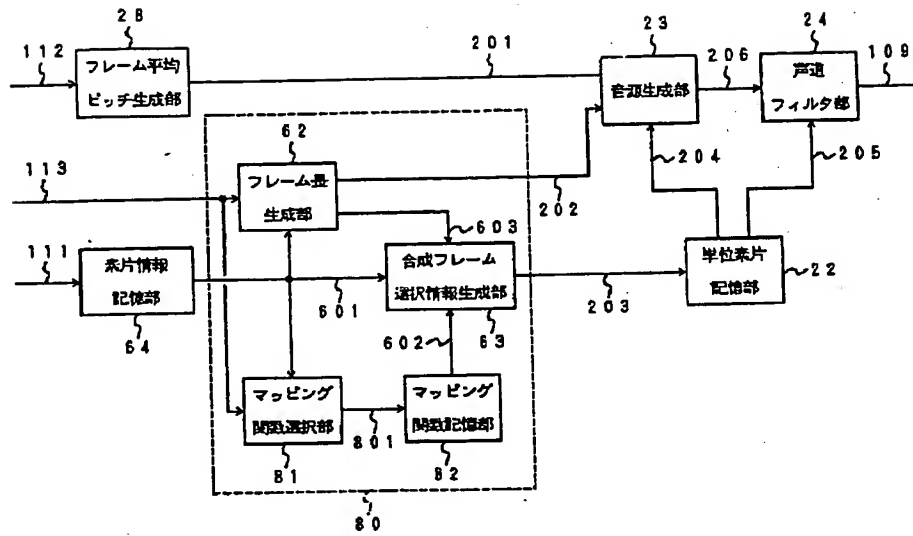
【図6】



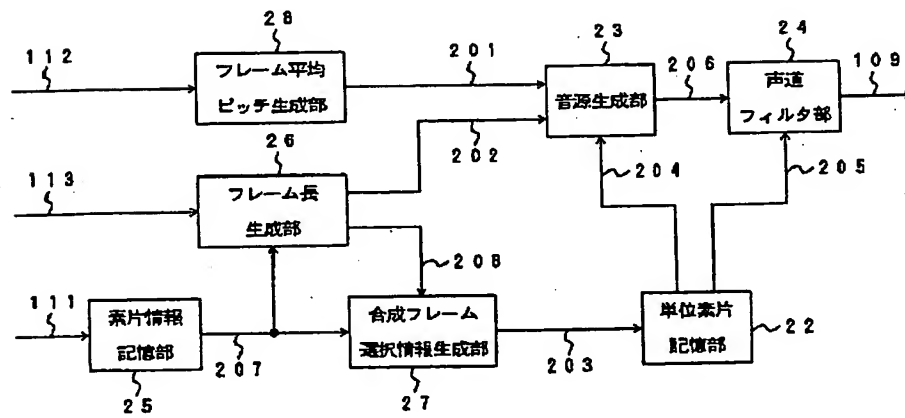
【図8】



【図7】



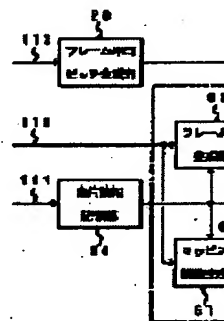
【図9】



Patent number: JP10091191
Publication date: 1998-04-10
Inventor: KAGOSHIMA TAKEHIKO, AKAMINE MASAMI
Applicant: TOSHIBA CORP
Classification:
- international: G10L5/04
- european:
Application number: JP19960246718 19960918
Priority number(s):

PROBLEM TO BE SOLVED: To provide a method of voice synthesis permitting to generate a synthesized voice for an arbitrary phoneme duration without causing damage to the nature.

SOLUTION: In a method of a voice synthesis for synthesizing voice signals 10 by using synthesis parameters (sound source parameter 204, voice path parameter 205) of a first frame unit representing the nature of voice signals to be synthesized, synthesis parameters are generated by storing unit element pieces of a second frame unit to become synthesis parameters in a unit element piece storage part 22 beforehand, giving a synthesis frame selection information generation part 63 a mapping function generated by a function generation part 61 according to information of the inputted phoneme duration 113, making the first and second frames corresponded to each other, and further, thinning out or repeating the unit element pieces of the second frame corresponding to the first frame according to the mapping function.



1 / 1

THIS PAGE BLANK (USPTO)